

A Sandwich Likelihood Correction for Bayesian Quantile Regression based on the Misspecified Asymmetric Laplace Density

KARTHIK SRIRAM*

Last revised on June 11, 2015

Abstract

A sandwich likelihood correction is proposed to remedy an inferential limitation of the Bayesian quantile regression approach based on the misspecified asymmetric Laplace density, by leveraging the benefits of the approach. Supporting theoretical results and simulations are presented.

Key words and phrases. Bayesian; Asymmetric Laplace; Credible Interval, Misspecification.

1 Introduction

Quantile Regression is used to model conditional quantiles of the independently distributed responses $\{Y_i\}_{i=1}^n$ given p -dimensional covariate vectors $\{\mathbf{X}_i\}_{i=1}^n$. The frequentist approach to modeling the τ^{th} quantile ($0 < \tau < 1$) as proposed by Koenker and Bassett (1978) involves solving the problem:

$$\beta_n^M = \arg \min_{\beta} \sum_{i=1}^n \rho_{\tau}(Y_i - \mathbf{X}_i^T \beta), \quad (1)$$

where $\rho_{\tau}(u) = u(\tau - I_{(u \leq 0)})$ with $I_{(\cdot)}$ being the indicator function. This procedure is equivalent to Maximum Likelihood Estimation (MLE) if the responses are assumed to follow the asymmetric Laplace distribution (ALD), whose probability density function (p.d.f) is given by

$$f_{i\beta}(y) = \tau(1 - \tau) \exp \{-\rho_{\tau}(y - \mu_i^{\tau})\}, \quad \text{with } \mu_i^{\tau} = \mathbf{X}_i^T \beta \text{ and } y \in (-\infty, \infty). \quad (2)$$

*Karthik Sriram is Assistant Professor, Production and Quantitative Methods Area, Indian Institute of Management Ahmedabad, India (karthiks@iimahd.ernet.in)

It is easy to check that μ_i^τ is the τ^{th} quantile with respect to the p.d.f $f_{i\beta}(\cdot)$.

The focus of this paper is a widely used Bayesian approach to the problem proposed by Yu and Moyeed (2001), where the Bayesian posterior is obtained by assuming the likelihood $f_{i\beta}(\cdot)$ for Y_i , and a prior $\Pi(\cdot)$ for β . The approach is computationally attractive especially due to the location scale mixture normal representation of ALD (see Kozumi and Kobayashi (2011)), and is known to give posterior consistent estimates even if ALD is a misspecification (see Sriram et al. 2013). Therefore, the approach has been useful in many applications (e.g. Yu et al. 2005; Yue and Rue 2011; Benoit and Van den Poel 2012; Alhamzawi and Yu 2013; Waldmann et al. 2013).

The aim of this paper is to highlight and remedy an inferential limitation of the Bayesian quantile regression approach based on ALD, when the ALD model is possibly misspecified. From a classical Bayesian point of view, where the objective is to draw inference on the “true” underlying quantile regression parameters (β_0), it is desirable that the Bayesian inference coincides with the frequentist inference as the size of data increases. For this, two “frequentist” asymptotic properties need to hold, (i) Posterior consistency, i.e. the posterior asymptotically concentrates around the “true” parameter value, and (ii) “Coverage property”, i.e. the $100(1-\alpha)\%$ Bayesian credible sets asymptotically merge with the $100(1-\alpha)\%$ frequentist confidence sets around β_n^M . Even from a subjectivist Bayesian point of view, which does not subscribe to the idea of a “true parameter” value, Diaconis and Freedman (1986) argue that violation of posterior consistency is undesirable. Such violation would mean that two experts starting with different priors, can completely diverge on their opinions about the predictive distributions, even as more data becomes available. Further, any small sample posterior inference based on a possibly misspecified likelihood is not straight forward to justify. In such cases, checking the large sample (asymptotic) properties is a way to avoid using an inappropriate likelihood.

The ALD is mainly used as a “working likelihood” for Bayesian quantile inference. It is more often than not a misspecification of the true underlying likelihood. While posterior consistency still holds under suitable conditions (see Sriram et al. 2013), we find that the “coverage property” may not hold. An undesirable consequence would be that a narrow Bayesian credible interval could give a false sense of certainty about a parameter, which is an artifact of a misspecified likelihood rather than an actual gain from the Bayesian approach. This paper proposes a sandwich likelihood method to remedy this issue, while still leveraging the benefits of using ALD.

Section 2 describes the sandwich likelihood method. Supporting theoretical results and simulations are presented in Sections 3 and 4 respectively.

2 The Sandwich Likelihood Method

Let $\tau \in (0, 1)$ be fixed. Suppose $\{Y_i, i = 1, 2, \dots, n\}$ are independent but non-identically distributed (*i.n.i.d*) with probability density function $Y_i \sim p_i(\cdot)$. Let the “true” τ^{th} quantile of Y_i be given by $Q_\tau(\mathbf{X}_i) = \mathbf{X}_i^T \boldsymbol{\beta}_0$, where \mathbf{X}_i is a vector of p -dimensional non-random covariates. We model the τ^{th} quantile as $Q_\tau(\mathbf{X}_i) = \mathbf{X}_i^T \boldsymbol{\beta}$ along with a proper prior Π (with p.d.f π) for $\boldsymbol{\beta}$. We write the posterior distribution of $\boldsymbol{\beta}$ given the data Y_1, Y_2, \dots, Y_n as

$$\Pi_n(\boldsymbol{\beta} \in B) = \frac{\int_B f_{\boldsymbol{\beta}}^{(n)} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}}{\int_{\mathcal{R}^d} f_{\boldsymbol{\beta}}^{(n)} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}}, \text{ where } f_{\boldsymbol{\beta}}^{(n)} = \prod_{i=1}^n f_{i\boldsymbol{\beta}}(Y_i). \quad (3)$$

The sandwich likelihood method is motivated from Müller (2013) and is based on two observations.

- (a) As seen later in Theorem 1, the posterior distribution $\Pi_n(\cdot)$ of $\boldsymbol{\beta}$ is asymptotically equivalent to a normal distribution centred at $\boldsymbol{\beta}_n^M$ (as in equation(1)) and covariance matrix $\frac{1}{n}V^{-1}$, where

$$V = \left(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n p_i(\mathbf{X}_i^T \boldsymbol{\beta}_0) \mathbf{X}_i \mathbf{X}_i^T \right). \quad (4)$$

- (b) It follows from Koenker 2005 (page 74) that $\boldsymbol{\beta}_n^M$ is asymptotically normal with mean $\boldsymbol{\beta}_0$ and (“sandwich”) covariance matrix $\frac{1}{n}\Sigma$, where

$$\Sigma = \tau(1 - \tau)V^{-1}SV^{-1} \text{ with } S = \lim_{n \rightarrow \infty} S_n, \quad S_n = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T. \quad (5)$$

Since $V^{-1} \neq \Sigma$ in general, (a) and (b) imply that the credible intervals from $\Pi_n(\cdot)$ will not asymptotically match the normal frequentist confidence intervals, thus violating the “coverage property” described in the introduction.

Let $\tilde{\boldsymbol{\beta}}_n$ and $\frac{1}{n}V_n^{-1}$ be the mean and covariance matrix for $\boldsymbol{\beta}$ under the posterior distribution $\Pi_n(\cdot)$. The “Sandwich likelihood” method to remedy this issue can be described in two steps.

Step 1.(ALD step) Using the posterior distribution $\Pi_n(\cdot)$ purely based on the ALD likelihood $Y_i \sim f_{i\boldsymbol{\beta}}$ and prior $\Pi(\cdot)$, compute the posterior mean $= \tilde{\boldsymbol{\beta}}_n$, and covariance matrix $= \frac{1}{n}V_n^{-1}$.

Step 2. (Sandwich Likelihood step) Define the “Sandwich likelihood” as

$$g_{\boldsymbol{\beta}}^{(n)} = \frac{1}{|\Sigma_n/n|^{\frac{1}{2}}} \cdot e^{-n(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^T \Sigma_n^{-1} (\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta})} \quad (6)$$

$$\text{where } \Sigma_n = \tau(1 - \tau)V_n^{-1}S_nV_n^{-1}, \quad (7)$$

and recompute a new posterior distribution for $\boldsymbol{\beta}$ as follows:

$$\Pi_n^{prop}(\boldsymbol{\beta} \in B) = \frac{\int_B g_{\boldsymbol{\beta}}^{(n)} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}}{\int_{\mathcal{R}^d} g_{\boldsymbol{\beta}}^{(n)} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta}}. \quad (8)$$

Theorem 2 in Section 3 ensures that the credible sets based on the new posterior from Step 2 merge asymptotically with the frequentist confidence sets. Here, we make a few remarks.

Remark 1. Note that the proposed sandwich likelihood is based on the Bayesian posterior mean $\tilde{\boldsymbol{\beta}}_n$ from the ALD step (i.e. Step 1). We denote this proposed approach by “SLBA”. An alternative approach to sandwich likelihood in Step 2 is in the lines of Müller (2013), got by using the classical quantile regression estimator $\boldsymbol{\beta}_n^M$ instead of $\tilde{\boldsymbol{\beta}}_n$ in equation (6). We will denote this alternative by “SLQR”. While both these methods are asymptotically equivalent (see Lemma 2a), and work similarly for relatively flat priors or large sample sizes, simulations in Section 4 suggest that the proposed SLBA method may be more suited for small sample sizes with informative priors.

Remark 2. The Markov Chain Monte Carlo (MCMC) approaches for implementing Step 1 are now well known (e.g. see Yue and Rue 2011). Therefore, $\tilde{\boldsymbol{\beta}}_n$ and $\frac{1}{n}V_n^{-1}$ can be obtained as the mean and covariance matrix computed based on the MCMC simulations of $\boldsymbol{\beta}$ from Step 1. Consequently, Σ_n (as in equation 7) needed for Step 2 can be easily computed.

Remark 3. Under the assumptions made in Section 3, Lemma 2 shows that $\tilde{\boldsymbol{\beta}}_n$ and V_n^{-1} are consistent for $\boldsymbol{\beta}_0$ and V^{-1} respectively. Consequently, Σ_n is consistent for Σ . It is worth noting that estimation of the Σ in the *i.n.i.d.* case is in general a challenging problem (see section 3.4.2 of Koenker 2005) since the true underlying densities $\{p_i(\cdot)\}_{i \geq 1}$ are not known. This method is a simple alternative.

3 Theoretical Results

Recall that the true underlying p.d.f of Y_i is p_i . By way of notation, let $P^{(n)}$ denote the product probability $p_1 \times p_2 \times \cdots p_n$ and P denote the infinite product probability $\prod_{i=1}^{\infty} p_i$. Let $P(\cdot)$ and

$E[\cdot]$ denote the probability and expectation with respect to the true product probability. We define $Z_i := Y_i - \mathbf{X}_i^T \boldsymbol{\beta}_0$ and note that $P(Z_i \leq 0 | \mathbf{X}_i) = \tau$.

Our Assumptions 1a, 2 to 4 are exactly those in Sriram et al. (2013) for posterior consistency. Assumption 1b is additionally introduced to help ensure consistency of posterior mean and variance.

Assumption 1.

- (a) Π is a proper prior with a bounded and continuous p.d.f π , with $\pi(\boldsymbol{\beta}_0) > 0$.
- (b) $\int \|\boldsymbol{\beta}\|^2 \pi(\boldsymbol{\beta}) d\boldsymbol{\beta} < \infty$.

The second assumption requires that the covariates be bounded.

Assumption 2. $\exists M > 0$ such that $\|\mathbf{X}_i\| \leq M, \forall i$.

The first part of the next assumption essentially says that the non-intercept covariates (after appropriate centering) take values in all quadrants of the Euclidean plane. In particular, this implies that they cannot be collinear. The second part of the assumption requires that the true underlying likelihood put positive mass around the true quantile, in particular ensuring that it is unique.

Assumption 3.

- (a) Let the first coordinate of \mathbf{X}_i be identically 1 representing the intercept. After appropriate centering of the other co-ordinates, $\exists \epsilon_0 > 0$ such that $\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n I_{\mathbf{X}_i \in D} > 0, \forall D \subset \mathcal{R}^p$ of the form $\{1\} \times U_2 \times \cdots \times U_p$, where for some j , U_j is either (ϵ_0, ∞) or $(-\infty, -\epsilon_0)$ and for $k \neq j$, U_k is either $(0, \infty)$ or $(-\infty, 0)$.
- (b) For some $C > 0$ and all sufficiently small $\Delta > 0$, $P(0 < Z_i < \Delta) > C\Delta$ and $P(-\Delta < Z_i < 0) > C\Delta \forall i$.

Assumption 4 is a technical condition required for Strong Law of Large Numbers in the *i.n.i.d* case.

Assumption 4. $\limsup_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m E|Z_i| < \infty$ and $\sum_{i=1}^{\infty} \frac{EZ_i^2}{i^2} < \infty$.

Lemma 1. *Let Assumptions 1 to 4 hold, and $k \in \{0, 1, 2\}$. Then, for any sequence $M_n \rightarrow \infty$,*

$$E \left[\int_{\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| > \frac{M_n}{\sqrt{n}}} (\sqrt{n} \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\|)^k d\Pi_n \right] \rightarrow 0.$$

For $k = 0$, Lemma 1 gives posterior consistency at \sqrt{n} -rate (the same conclusion as in Theorem 2b of Sriram et al. (2013)). The results for $k = 1$ and $k = 2$ are useful in establishing the next lemma, which formalizes the fact that the classical estimator (β_n^M) and the posterior mean from Step 1 $(\tilde{\beta}_n)$ are asymptotically equivalent, and that the estimators (V_n^{-1}, Σ_n) are consistent for (V^{-1}, Σ) .

Lemma 2. *Under Assumption 1 to 4,*

- (a) $\sqrt{n}(\beta_n^M - \tilde{\beta}_n) \rightarrow 0$ in probability $[P]$.
- (b) $V_n^{-1} \rightarrow V^{-1}$, and hence $\Sigma_n \rightarrow \Sigma$ in probability $[P]$.

The asymptotic normality of the posterior is derived by applying the Bernstein-von-Mises theorem for misspecified models given by Kleijn and van der Vaart (2012). For that, a key requirement apart from \sqrt{n} -posterior consistency, is “Local Asymptotic Normality” (LAN). For the ALD model on i.n.i.d data, the LAN property can be shown by making the following assumption on the boundedness and continuity of the true underlying densities p_i .

Assumption 5.

For some $C, \eta > 0$ and for all β in some small enough neighborhood of β_0 , the p.d.fs p_i satisfy:

- (a) $\{p_i(\mathbf{X}_i^T \beta_0), i \geq 1\}$ are uniformly bounded away from ∞ .
- (b) $|p_i(\mathbf{X}_i^T \beta) - p_i(\mathbf{X}_i^T \beta_0)| \leq C \|\mathbf{X}_i^T \beta - \mathbf{X}_i^T \beta_0\|^\eta \forall i$.

Lemma 3 (LAN property). *Under Assumptions 2 and 5, the LAN property holds, i.e., for any compact set $K \subset \mathcal{R}^p$,*

$$\sup_{\delta \in K} \left| \log \frac{f_{\beta_0 + \frac{\delta}{\sqrt{n}}}^{(n)}}{f_{\beta_0}^{(n)}} - \delta^T V \Delta_{n, \beta_0} - \frac{1}{2} \delta^T V \delta \right| \rightarrow 0 \text{ in probability } [P]. \quad (9)$$

$$\text{where } \Delta_{n, \beta_0} = -V^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\tau - I_{(Y_i \leq \mathbf{X}_i^T \beta_0)}) \mathbf{X}_i. \quad (10)$$

Proofs of Lemmas 1, 2 and 3 are included in the Appendix and are essentially extensions of ideas from Sriram et al. (2013), Kleijn and van der Vaart (2012) and Koenker (2005) respectively. The next lemma establishes the asymptotic connection between the posterior probability and the normal distribution. Let $\Phi(B, \mu, \Sigma)$ denote the probability of a set B under the multivariate normal

distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then, we have the following result.

Lemma 4. *Under Assumptions 1 to 5,*

$$(a) \quad \sup_B \left| \Pi_n \left(\sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \in B \right) - \Phi \left(B, \Delta_{n, \boldsymbol{\beta}_0}, V^{-1} \right) \right| \rightarrow 0 \text{ in probability } [P]. \quad (11)$$

$$(b) \quad \sqrt{n}(\boldsymbol{\beta}_n^M - \boldsymbol{\beta}_0) - \Delta_{n, \boldsymbol{\beta}_0} \rightarrow 0 \text{ in probability } [P]. \quad (12)$$

Proof. Part(a) is an immediate consequence of Lemma 1 (for $k = 0$) and Lemma 3, since these are precisely the conditions needed to apply Theorem 2.1 of Kleijn and van der Vaart (2012). Part (b) is shown in Koenker (2005) (see page 122, equation 4.4). \square

Our first theorem is immediate from Lemma 4 and formalizes Step 1.

Theorem 1. *Let Π_n be as in equations (3). Then, under assumptions 1 to 5,*

$$\sup_B \left| \Pi_n \left(\sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \in B \right) - \Phi \left(B, \sqrt{n}(\boldsymbol{\beta}_n^M - \boldsymbol{\beta}_0), V^{-1} \right) \right| \rightarrow 0 \text{ in probability } [P].$$

The next theorem formalizes Step 2 of the proposed Sandwich Likelihood approach.

Theorem 2. *Let Π_n and Π_n^{prop} be as in equations (3) and (8). Then, under assumptions 1 to 5,*

$$(a) \quad \sup_B \left| \Pi_n \left(\sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \in B \right) - \Phi \left(B, \sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0), V^{-1} \right) \right| \rightarrow 0 \text{ in probability } [P].$$

$$(b) \quad \sup_B \left| \Pi_n^{prop} \left(\sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \in B \right) - \Phi \left(B, \sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0), \boldsymbol{\Sigma} \right) \right| \rightarrow 0 \text{ in probability } [P].$$

Proof. Part (a) follows from Theorem 1 and Lemma 2 (a). To see part (b), first note that $\boldsymbol{\Sigma}_n \rightarrow \boldsymbol{\Sigma}$ in probability $[P]$ (by Lemma 2). Since $\sqrt{n}(\boldsymbol{\beta}_n^M - \boldsymbol{\beta}_0)$ is asymptotically normal (Koenker (2005)), it is bounded in probability. Then, by Lemma 2 (a), $\sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ is also bounded in probability. Hence, $g_{\boldsymbol{\beta}}^{(n)}$ satisfies LAN property because

$$\begin{aligned} & \sup_{\boldsymbol{\delta} \in \mathcal{K}} \left| \log \frac{g_{\boldsymbol{\beta}_0 + \frac{\boldsymbol{\delta}}{\sqrt{n}}}^{(n)}}{g_{\boldsymbol{\beta}_0}^{(n)}} + \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) - \frac{1}{2} \boldsymbol{\delta}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\delta} \right|. \\ &= \sup_{\boldsymbol{\delta} \in \mathcal{K}} \left| \boldsymbol{\delta}^T (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}_n^{-1}) \sqrt{n}(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) - \frac{1}{2} \boldsymbol{\delta}^T (\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}_n^{-1}) \boldsymbol{\delta} \right| \rightarrow 0 \text{ in probability } [P]. \end{aligned}$$

The result is immediate from Theorem 2.1 of Kleijn and van der Vaart (2012), provided we show

$$E [\Pi_n^{prop} (\sqrt{n} \|\beta - \beta_0\| > M_n)] \rightarrow 0 \text{ for any sequence } M_n \rightarrow \infty. \quad (13)$$

Let $B_n = \{\beta : \sqrt{n} \|\beta - \beta_0\| > M_n\}$. Making a change of variable $\sqrt{n}(\beta - \tilde{\beta}_n) = \mathbf{t}$, we can write

$$\Pi_n^{prop}(B_n) = \frac{\int_{\mathcal{R}^d} \left(\frac{1}{\sqrt{2\pi \cdot \det(\Sigma)}} \right)^d \cdot e^{-\frac{\mathbf{t}^T \Sigma^{-1} \mathbf{t}}{2}} \cdot I_{(\|\tilde{\beta}_n + \mathbf{t}/\sqrt{n} - \beta_0\| > M_n)} \cdot \pi(\tilde{\beta}_n + \mathbf{t}/\sqrt{n}) d\mathbf{t}}{\int_{\mathcal{R}^d} \left(\frac{1}{\sqrt{2\pi \cdot \det(\Sigma)}} \right)^d \cdot e^{-\frac{\mathbf{t}^T \Sigma^{-1} \mathbf{t}}{2}} \cdot \pi(\tilde{\beta}_n + \mathbf{t}/\sqrt{n}) d\mathbf{t}}$$

Since $\tilde{\beta}_n \rightarrow \beta_0$ in probability $[P]$, $M_n \rightarrow \infty$, and $\pi(\cdot)$ is bounded and continuous by Assumption 1, an application of Skorohod representation theorem for the sequence $\{\tilde{\beta}_n\}_{n \geq 1}$ along with the dominated convergence theorem implies that the numerator of the above expression converges to zero and the denominator to $\pi(\beta_0)$. This in turn implies that $E [\Pi_n^{prop} (\sqrt{n} \|\beta - \beta_0\| > M_n)] \rightarrow 0$. \square

Remark 4 (On using ALD with a scale parameter). In some applications, one may carry out Bayesian quantile regression using an ALD that includes a scale parameter ($\sigma > 0$), whose p.d.f is given by

$$f_{i\beta, \sigma}(y) = \frac{\tau(1-\tau)}{\sigma} \exp \left\{ -\frac{\rho_\tau(y - \mu_i^\tau)}{\sigma} \right\}, \text{ with } \mu_i^\tau = \mathbf{X}_i^T \beta \text{ and } y \in (-\infty, \infty). \quad (14)$$

Here, either the scale is fixed (e.g. $\sigma = 1$) or endowed with a prior. In the latter case, the scale is essentially a nuisance parameter for Bayesian inference on β . The proposed approach can be easily modified to incorporate the scale parameter. Suppose the scale is fixed at $\sigma = \sigma_0$. Then under Assumptions 1 to 5, it is easy to check that the conclusions of Theorem 2(a) will hold with V^{-1} replaced by $\sigma_0 V^{-1}$. Accordingly, $\frac{1}{n} V_n^{-1}$ in Step 1 can be obtained as $\frac{1}{\sigma_0}$ times the estimated covariance matrix under Π_n . Step 2 would remain the same.

Suppose σ is endowed with a prior on a compact interval $[\sigma_1, \sigma_2]$. Then, Sriram et al. (2013) show under suitable conditions, that the posterior distribution of σ given Y_1, Y_2, \dots, Y_n would concentrate around a value σ_0 given by

$$\sigma_0 = \arg \max_{\sigma \in [\sigma_1, \sigma_2]} \log \left(\frac{\tau(1-\tau)}{\sigma} \right) - \frac{C^*}{\sigma}, \quad (15)$$

$$\text{where, } C^* = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m E(Z_i(\tau - I_{(Z_i \leq 0)})) \text{ and } Z_i = Y_i - \mathbf{X}_i \beta_0.$$

In this case, it is reasonable to expect that under suitable conditions, conclusions of Theorem 2(a) will hold with V^{-1} replaced by $\sigma_0 V^{-1}$, only now with σ_0 as in equation (15). It appears that a formal derivation of this result requires the Bernstein-von-Mises theorem for misspecified models (as in Kleijn and van der Vaart 2012) to be developed in the presence of a nuisance parameter. Hence, a formal investigation is deferred to a future work. Here, one could first estimate σ_0 from Step 1 using the MCMC simulations of the parameter σ (e.g. posterior mean $\hat{\sigma}_0$), and then obtain $\frac{1}{n} V_n^{-1}$ as $\frac{1}{\hat{\sigma}_0}$ times the covariance matrix under Π_n .

Table 1: Comparison of methods when data size is large (N=2000). The numbers within parenthesis are coverage and interval length (COV in %, LEN).

(a): N=2000, relatively flat prior on β , Fixed ALD scale parameter at =1.

Model	τ	α				β_1				β_2			
		QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA
1	0.25	(94.0.43)	(98.0.55)	(95.0.42)	(96.0.42)	(94.0.13)	(98.0.17)	(94.0.13)	(95.0.13)	(94.0.26)	(98.0.34)	(93.0.26)	(92.0.26)
2	0.25	(92.0.18)	(100.0.36)	(92.0.18)	(94.0.18)	(94.0.06)	(100.0.11)	(94.0.06)	(95.0.06)	(94.0.11)	(100.0.22)	(94.0.11)	(96.0.11)
3	0.25	(96.2.54)	(72.1.31)	(96.2.52)	(95.2.52)	(98.0.92)	(64.0.44)	(94.0.92)	(94.0.92)	(96.2.17)	(68.0.96)	(96.2.15)	(96.2.15)
4	0.25	(95.2.76)	(72.1.36)	(95.2.72)	(95.2.72)	(95.1)	(68.0.46)	(96.0.99)	(97.0.99)	(94.2.4)	(57.1)	(92.2.33)	(92.2.33)
1	0.75	(94.0.42)	(98.0.53)	(94.0.41)	(94.0.41)	(94.0.13)	(98.0.16)	(92.0.13)	(94.0.13)	(94.0.26)	(100.0.34)	(93.0.26)	(94.0.26)
2	0.75	(94.0.52)	(98.0.6)	(91.0.51)	(92.0.51)	(93.0.16)	(96.0.18)	(92.0.16)	(92.0.16)	(96.0.33)	(98.0.38)	(94.0.33)	(96.0.33)
3	0.75	(95.1.75)	(83.1.07)	(94.1.72)	(94.1.72)	(97.0.64)	(78.0.36)	(96.0.63)	(96.0.63)	(94.1.51)	(70.0.79)	(92.1.48)	(92.1.48)
4	0.75	(94.2.68)	(66.1.31)	(94.2.62)	(92.2.62)	(96.0.97)	(64.0.44)	(94.0.95)	(94.0.95)	(96.2.27)	(68.0.96)	(94.2.19)	(94.2.19)

(b): N=2000, relatively flat prior on β , Inverse gamma prior on ALD scale parameter.

Model	τ	α				β_1				β_2			
		QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA
1	0.25	(94.0.41)	(86.0.3)	(95.0.41)	(94.0.41)	(96.0.13)	(86.0.09)	(96.0.12)	(95.0.12)	(98.0.26)	(83.0.19)	(96.0.26)	(96.0.26)
2	0.25	(96.0.18)	(96.0.16)	(96.0.18)	(97.0.18)	(97.0.05)	(96.0.05)	(96.0.05)	(97.0.05)	(91.0.11)	(90.0.1)	(91.0.11)	(92.0.11)
3	0.25	(96.2.44)	(96.2.17)	(96.2.4)	(98.2.4)	(96.0.88)	(92.0.72)	(96.0.87)	(97.0.87)	(93.2.11)	(88.1.62)	(92.2.08)	(94.2.08)
4	0.25	(94.2.71)	(86.2.11)	(91.2.63)	(94.2.63)	(94.0.98)	(86.0.7)	(92.0.96)	(92.0.96)	(95.2.36)	(80.1.59)	(92.2.32)	(92.2.32)
1	0.75	(95.0.42)	(84.0.3)	(93.0.41)	(94.0.41)	(96.0.13)	(82.0.09)	(94.0.13)	(94.0.13)	(95.0.27)	(84.0.19)	(96.0.27)	(96.0.27)
2	0.75	(96.0.53)	(80.0.35)	(94.0.52)	(94.0.52)	(96.0.16)	(80.0.11)	(94.0.16)	(93.0.16)	(95.0.34)	(80.0.22)	(92.0.33)	(94.0.33)
3	0.75	(92.1.68)	(80.1.25)	(90.1.65)	(90.1.65)	(89.0.62)	(76.0.42)	(89.0.61)	(88.0.61)	(94.1.54)	(78.0.97)	(93.1.5)	(94.1.5)
4	0.75	(94.2.63)	(91.2.07)	(94.2.58)	(94.2.58)	(96.0.98)	(84.0.7)	(94.0.96)	(94.0.96)	(92.2.36)	(80.1.59)	(90.2.33)	(90.2.33)

(c): N=2000, Informative prior on β , Fixed ALD scale parameter at =1.

Model	τ	α				β_1				β_2			
		QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA
1	0.25	(96.0.42)	(100.0.53)	(98.0.41)	(96.0.41)	(96.0.13)	(100.0.16)	(96.0.13)	(96.0.13)	(96.0.26)	(100.0.34)	(95.0.26)	(96.0.26)
2	0.25	(94.0.18)	(100.0.35)	(95.0.17)	(98.0.17)	(96.0.06)	(100.0.11)	(96.0.05)	(96.0.05)	(95.0.11)	(100.0.22)	(96.0.11)	(95.0.11)
3	0.25	(98.2.43)	(78.1.19)	(98.1.82)	(99.1.82)	(96.0.88)	(69.0.4)	(96.0.69)	(97.0.69)	(96.2.12)	(68.0.91)	(96.1.72)	(96.1.72)
4	0.25	(96.2.68)	(76.1.24)	(96.1.94)	(98.1.94)	(97.0.98)	(72.0.42)	(95.0.74)	(96.0.74)	(94.2.36)	(68.0.95)	(92.1.82)	(96.1.82)
1	0.75	(94.0.42)	(98.0.53)	(93.0.4)	(95.0.4)	(94.0.13)	(98.0.16)	(92.0.12)	(94.0.12)	(94.0.26)	(100.0.34)	(93.0.26)	(93.0.26)
2	0.75	(94.0.52)	(98.0.59)	(90.0.49)	(92.0.49)	(93.0.16)	(96.0.18)	(91.0.15)	(94.0.15)	(96.0.33)	(98.0.38)	(95.0.33)	(97.0.33)
3	0.75	(95.1.75)	(84.1.02)	(94.1.45)	(95.1.45)	(97.0.64)	(78.0.34)	(96.0.54)	(96.0.54)	(94.1.51)	(72.0.78)	(94.1.33)	(94.1.33)
4	0.75	(94.2.68)	(70.1.24)	(94.1.93)	(96.1.93)	(96.0.97)	(66.0.41)	(93.0.74)	(95.0.74)	(96.2.27)	(68.0.93)	(96.1.78)	(96.1.78)

(d): N=2000, Informative prior on β , Inverse gamma prior on ALD scale parameter.

Model	τ	α				β_1				β_2			
		QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA
1	0.25	(96.0.42)	(86.0.29)	(95.0.4)	(96.0.4)	(96.0.13)	(88.0.09)	(94.0.12)	(96.0.12)	(94.0.27)	(86.0.19)	(94.0.26)	(94.0.26)
2	0.25	(96.0.17)	(95.0.16)	(96.0.17)	(94.0.17)	(94.0.05)	(92.0.05)	(94.0.05)	(93.0.05)	(94.0.12)	(94.0.1)	(94.0.11)	(94.0.11)
3	0.25	(94.2.39)	(96.1.85)	(88.1.59)	(97.1.59)	(95.0.88)	(94.0.63)	(90.0.62)	(96.0.62)	(96.2.2)	(94.1.53)	(94.1.64)	(98.1.64)
4	0.25	(94.2.71)	(91.1.84)	(88.1.76)	(98.1.76)	(92.0.98)	(86.0.62)	(90.0.68)	(96.0.68)	(95.2.44)	(85.1.51)	(93.1.81)	(96.1.81)
1	0.75	(95.0.42)	(88.0.3)	(94.0.41)	(94.0.41)	(96.0.13)	(87.0.09)	(96.0.13)	(96.0.13)	(92.0.27)	(86.0.19)	(92.0.26)	(92.0.26)
2	0.75	(96.0.54)	(82.0.36)	(96.0.52)	(96.0.52)	(96.0.16)	(84.0.11)	(97.0.16)	(97.0.16)	(96.0.34)	(84.0.22)	(96.0.34)	(96.0.34)
3	0.75	(98.1.73)	(91.1.22)	(94.1.4)	(98.1.4)	(96.0.63)	(84.0.41)	(92.0.52)	(96.0.52)	(94.1.56)	(78.0.95)	(90.1.33)	(92.1.33)
4	0.75	(94.2.73)	(94.1.86)	(89.1.75)	(96.1.75)	(94.0.99)	(88.0.63)	(89.0.69)	(98.0.69)	(94.2.38)	(84.1.47)	(94.1.74)	(98.1.74)

4 Simulation Study

Here, we study the performance of the proposed sandwich likelihood method (SLBA) by simulating data from different “true” underlying models. We work with two covariates X_1 and X_2 , simulated from $N(3, 1)$ (truncated between 1 and 1000) and $Bernoulli(0.3)$ respectively. For each simulated model, we ensure that the true τ^{th} -quantile is given by $q_\tau(\mathbf{X}) = (1 + 2X_1 + 3X_2)$. The simulated models are described below. The first two models are a location shifted normal and a location shifted gamma respectively, the third is a scaled gamma, and the fourth model a location shifted and scaled normal.

Model 1. $Y = q_\tau(\mathbf{X}) + \epsilon$, where $\epsilon = Z - \rho_\tau$, $Z \sim N(0, 1)$ and $\rho_\tau = \tau^{th}$ quantile of $N(0, 1)$.

Model 2. $Y = 1 + 2X_1 + 3X_2 - \rho_\tau + e$, where $e \sim \text{Gamma}(\text{shape} = 1, \text{scale} = 1)$, ρ_τ is the τ^{th} quantile of $\text{Gamma}(1, 1)$

Model 3. $Y \sim \text{Gamma}\left(\text{shape} = 2, \text{scale} = \frac{\rho_\tau}{q_\tau(\mathbf{X})}\right)$, where $\rho_\tau = \tau^{th}$ quantile of $\text{Gamma}(2, 1)$.

Model 4. $Y \sim N(1 + 2X_1 + 3X_2 - \rho_\tau | 1 + 2X_1 + 3X_2|, |1 + 2X_1 + 3X_2|^2)$, where ρ_τ is the τ^{th} quantile of $N(0, 1)$.

The specified model for the τ^{th} quantile is $Q_\tau(\mathbf{X}) = \alpha + \beta_1 X_1 + \beta_2 X_2$. We present results for different quantiles ($\tau \in \{0.25, .75\}$), different sample sizes ($N \in \{50, 2000\}$) and analyze with respect to a relatively flat prior (i.e. a product of three $N(0, 100)$ distributions), as well as an informative prior for the parameters $(\alpha, \beta_1, \beta_2)$ (i.e., a product of $N(.9, 1)$, $N(2.1, 1)$ and $N(2.9, 1)$). In addition, we present two scenarios for the scale parameter of ALD, (i) $\sigma = 1$ and (ii) $\sigma \sim \text{Gamma}$ prior with mean=100 and variance=1000. The conclusions reached were similar for other scenarios involving values of $\tau \in \{0.05, .5, .95\}$, $\sigma \in \{1/2, 2\}$, and $N \in \{100, 500\}$. However, these scenarios are not shown here in the interest of conciseness.

Recall that the proposed sandwich likelihood (denoted SLBA) is based on the posterior mean $\tilde{\beta}_n$ from Step 1. We compare this with the alternative method of using the classical estimator β_n^M (denoted by SLQR). We also compare with the frequentist quantile regression (denoted QR) and the Bayesian approach based purely on ALD, which is same as Step 1 (denoted ALD). For the QR method, we use the bootstrap method for computing confidence intervals and for the other methods we compute the posterior Bayesian credible intervals using 1000 MCMC simulations after a burn-in of 2000 simulations. The methods are compared with respect to the coverage property of the 95%

confidence/credible interval (denoted by COV), and the length of the confidence interval (denoted by LEN). The coverage (COV) is computed by repeating the above simulation 200 times and then calculating the percentage of times the 95% confidence/credible intervals contain the true value. Similarly, the length (LEN) is computed as the average interval length across the 200 repetitions.

Table 1 compares the methods across different scenarios when the sample size is large ($N = 2000$), and Table 2 does the comparison when sample size is small ($N = 50$). The sub-tables (a) and (b) are for a relatively flat prior on β , and sub-tables (c) and (d) are for an informative prior. Further, sub-table (a) and (c) fix the ALD scale parameter at 1, whereas sub-tables (b) and (d) carry out the analysis by considering a gamma prior on the ALD scale parameter.

We can make the following observations from Table 1. For large $N(=2000)$, it is desirable that Bayesian and the classical inferences merge asymptotically. Here, as expected from our results, the coverages and length of intervals from both the sandwich likelihood methods viz., the proposed SLBA and the alternative SLQR, as well as the classical QR methods are close to each other. SLBA is closer to SLQR for relatively flat priors, and performs slightly better with an informative prior. However, compared to these methods, the coverages from the ALD method in Tables 1(a) and (c) (i.e. scale=1) are not as close to 95%. For the simpler models 1 and 2 with *i.i.d.* errors, coverages under ALD are consistently but slightly higher than 95%. For more complex models 3 and 4 with *i.n.i.d* errors, coverages are way below 95%. So, for more complex models, the inadequacy of coverage under ALD is more pronounced. Such an issue with ALD can be partly addressed by allowing some flexibility in the ALD scale parameter. This is checked in Tables 1(b) and 1(d), where a prior is assumed on σ and the coverages from the ALD method are seen to improve for models 3 and 4. Since ALD is still a misspecification, the issue does not fully go away (e.g. as seen in the coverages for β_1 in Table 1(d) for $\tau = .75$). Finally, since the sample size is large, the results are similar for a relatively flat prior and for an informative prior. In summary, the simulation results in Table 1 are supportive of the asymptotic results in Section 3. They also highlight the fact that using Bayesian inference based purely on ALD could be more misleading especially when the true data generating likelihood is complex.

To further help demonstrate the usefulness of SLBA, Table 2 shows the simulation results for a small sample size $N = 50$. When the prior is relatively flat, as in sub-tables (a) and (b), the observations made in the previous paragraph still more or less hold. However, when we use an informative prior, as in sub tables (c) and (d), differences start to show. Unlike sub tables (a)

Table 2: Comparison of methods when data size is small (N=50). The numbers within parenthesis are coverage and interval length (COV in %, LEN).

(a): N=50, relatively flat prior on β , Fixed ALD scale parameter at =1.

Model	τ	α				β_1				β_2			
		QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA
1	0.25	(92,2.78)	(100,3.53)	(95,2.91)	(98,2.91)	(94,0.88)	(100,1.12)	(96,0.93)	(97,0.93)	(94,1.92)	(100,2.43)	(96,2.06)	(98,2.06)
2	0.25	(97,1.26)	(100,2.72)	(100,1.72)	(100,1.72)	(98,0.4)	(100,0.87)	(100,0.56)	(100,0.56)	(96,0.88)	(100,1.88)	(100,1.23)	(100,1.23)
3	0.25	(96,16.5)	(75,7.37)	(94,12.59)	(95,12.59)	(97,6.11)	(71,2.55)	(92,4.81)	(94,4.81)	(96,14.66)	(67,5.92)	(90,11.98)	(92,11.98)
4	0.25	(96,18.05)	(78,7.7)	(92,13.64)	(94,13.64)	(96,6.72)	(68,2.68)	(92,5.29)	(92,5.29)	(96,15.78)	(66,6.23)	(91,13.09)	(93,13.09)
1	0.75	(94,2.67)	(100,3.32)	(95,2.73)	(96,2.73)	(96,0.78)	(100,0.98)	(96,0.82)	(98,0.82)	(94,1.68)	(100,2.11)	(96,1.74)	(99,1.74)
2	0.75	(98,3.28)	(100,3.55)	(97,3.17)	(99,3.17)	(97,0.94)	(100,1.03)	(98,0.92)	(99,0.92)	(95,2.19)	(99,2.34)	(96,2.17)	(98,2.17)
3	0.75	(96,12.84)	(82,6.57)	(94,11.05)	(95,11.05)	(94,4.43)	(74,2.1)	(91,3.88)	(92,3.88)	(98,9.17)	(72,4.56)	(96,8.24)	(95,8.24)
4	0.75	(98,20.55)	(74,8.13)	(95,15.6)	(96,15.6)	(96,6.97)	(67,2.59)	(90,5.44)	(91,5.44)	(98,14.16)	(64,5.55)	(92,11.71)	(95,11.71)

(b): N=50, relatively flat prior on β , Inverse gamma prior on ALD scale parameter.

Model	τ	α				β_1				β_2			
		QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA
1	0.25	(97,3.37)	(80,2.01)	(93,3.25)	(92,3.25)	(96,1.05)	(80,0.62)	(92,1)	(92,1)	(97,1.82)	(80,1.09)	(95,1.72)	(96,1.72)
2	0.25	(98,1.49)	(94,1.12)	(96,1.44)	(98,1.44)	(99,0.47)	(96,0.35)	(97,0.45)	(97,0.45)	(96,0.84)	(92,0.61)	(93,0.77)	(97,0.77)
3	0.25	(98,21.96)	(94,14.06)	(93,15.35)	(97,15.35)	(98,7.43)	(94,4.6)	(93,5.34)	(97,5.34)	(96,14.71)	(86,8.98)	(92,11.86)	(96,11.86)
4	0.25	(98,22.92)	(92,13.3)	(96,16.2)	(98,16.2)	(98,7.74)	(86,4.35)	(94,5.67)	(96,5.67)	(94,15.44)	(86,8.7)	(93,12.94)	(94,12.94)
1	0.75	(94,3.01)	(84,1.82)	(91,2.9)	(93,2.9)	(94,0.94)	(83,0.56)	(90,0.92)	(93,0.92)	(96,2.16)	(90,1.28)	(96,2.08)	(96,2.08)
2	0.75	(94,3.63)	(80,2.05)	(89,3.5)	(92,3.5)	(93,1.11)	(82,0.63)	(91,1.08)	(92,1.08)	(94,2.64)	(82,1.42)	(90,2.44)	(93,2.44)
3	0.75	(93,12.81)	(86,7.57)	(92,10.97)	(94,10.97)	(94,4.67)	(82,2.55)	(92,4.06)	(94,4.06)	(93,13.49)	(73,6.38)	(90,10.93)	(92,10.93)
4	0.75	(96,19.67)	(88,12.13)	(91,15.08)	(96,15.08)	(92,7.01)	(80,4.04)	(90,5.47)	(93,5.47)	(94,20.65)	(80,10.44)	(91,15.54)	(92,15.54)

(c): N=50, Informative prior on β , Fixed ALD scale parameter at =1.

Model	τ	α				β_1				β_2			
		QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA
1	0.25	(96,2.69)	(100,2.46)	(84,1.38)	(98,1.38)	(96,0.79)	(100,0.76)	(86,0.48)	(98,0.48)	(96,1.7)	(100,1.79)	(94,1.19)	(98,1.19)
2	0.25	(96,1.18)	(100,2.04)	(98,0.99)	(100,0.99)	(96,0.35)	(100,0.63)	(98,0.33)	(100,0.33)	(95,0.75)	(100,1.4)	(98,0.74)	(99,0.74)
3	0.25	(98,18.89)	(98,3.35)	(36,2.28)	(100,2.28)	(96,6.39)	(79,1.31)	(74,1.61)	(92,1.61)	(96,12.68)	(87,3.09)	(66,2.74)	(98,2.74)
4	0.25	(96,19.72)	(98,3.37)	(30,2.3)	(99,2.3)	(92,6.71)	(78,1.33)	(72,1.69)	(91,1.69)	(96,13.81)	(86,3.14)	(64,2.79)	(98,2.79)
1	0.75	(94,2.67)	(100,2.45)	(78,1.38)	(98,1.38)	(96,0.78)	(100,0.76)	(84,0.48)	(98,0.48)	(94,1.68)	(100,1.8)	(90,1.2)	(98,1.2)
2	0.75	(98,3.28)	(100,2.55)	(82,1.48)	(100,1.48)	(97,0.94)	(100,0.79)	(86,0.52)	(100,0.52)	(95,2.19)	(100,1.93)	(91,1.37)	(98,1.37)
3	0.75	(96,12.84)	(96,3.23)	(44,2.17)	(97,2.17)	(94,4.43)	(88,1.23)	(76,1.41)	(92,1.41)	(98,9.17)	(90,2.92)	(72,2.57)	(98,2.57)
4	0.75	(98,20.55)	(98,3.4)	(34,2.32)	(98,2.32)	(96,6.97)	(77,1.34)	(72,1.69)	(88,1.69)	(98,14.16)	(88,3.15)	(65,2.79)	(100,2.79)

(d): N=50, Informative prior on β , Inverse gamma prior on ALD scale parameter.

Model	τ	α				β_1				β_2			
		QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA	QR	ALD	SLQR	SLBA
1	0.25	(94,2.68)	(86,1.48)	(92,1.82)	(95,1.82)	(95,0.78)	(84,0.43)	(92,0.56)	(94,0.56)	(96,1.71)	(87,0.98)	(94,1.36)	(95,1.36)
2	0.25	(98,1.15)	(94,0.85)	(95,0.97)	(98,0.97)	(98,0.34)	(94,0.25)	(96,0.29)	(96,0.29)	(96,0.76)	(93,0.55)	(92,0.66)	(96,0.66)
3	0.25	(98,19.18)	(100,3.6)	(12,1.27)	(98,1.27)	(95,6.5)	(98,1.67)	(52,1.54)	(94,1.54)	(98,13.5)	(100,3.57)	(28,1.87)	(98,1.87)
4	0.25	(98,19.58)	(100,3.57)	(17,1.44)	(100,1.44)	(96,6.7)	(94,1.63)	(54,1.62)	(94,1.62)	(96,14.52)	(100,3.52)	(34,2.05)	(98,2.05)
1	0.75	(92,2.7)	(81,1.5)	(92,1.84)	(92,1.84)	(94,0.82)	(82,0.46)	(90,0.59)	(96,0.59)	(96,1.77)	(82,1.03)	(94,1.41)	(96,1.41)
2	0.75	(94,3.34)	(85,1.67)	(93,2.02)	(96,2.02)	(95,1.02)	(86,0.52)	(94,0.68)	(96,0.68)	(96,2.31)	(84,1.2)	(95,1.68)	(96,1.68)
3	0.75	(96,11.39)	(100,3.27)	(38,1.94)	(100,1.94)	(94,4)	(89,1.27)	(66,1.31)	(92,1.31)	(94,10.29)	(96,3.15)	(68,2.46)	(100,2.46)
4	0.75	(96,18.73)	(100,3.56)	(16,1.42)	(98,1.42)	(95,6.66)	(90,1.62)	(51,1.61)	(92,1.61)	(98,15.68)	(100,3.58)	(32,2.01)	(99,2.01)

and (b), where the lengths of credible intervals for the QR and SLBA methods are similar, the SLBA method has much smaller intervals than QR in sub-tables (c) and (d), while still retaining coverages around 95%. This should not be surprising since the QR method is a classical approach and does not utilize the prior information, whereas the SLBA method utilizes the informative prior. In contrast, the SLQR method is seen to perform poorly. This is because SLQR leads to intervals of similar length as SLBA, but is centered at the classical estimator (β_n^M). Since β_n^M does not use the informative prior, this centering can be inaccurate. The ALD method performs better with an informative prior, but SLBA continues to have better coverage.

In summary, the proposed SLBA method works better than the SLQR, ALD and QR methods across all scenarios. Further, while SLBA and SLQR are asymptotically equivalent and perform

similarly in the case of large sample sizes or relatively flat priors, the SLBA method is seen to consistently work better and more pronouncedly so for small sample sizes and informative priors.

Appendix

Proof of Lemma 1. The proof is by extending ideas from Sriram et al. (2013). A sketch is provided here. Let $\Delta_n = \frac{M_n}{\sqrt{n}}$ and $k \in \{0, 1, 2\}$. Following arguments leading up to Lemma 5 of their paper, a compact set $G \subseteq \{\|\beta - \beta_0\| \leq M_0\}$ exists such that for some $u > 0$ and sufficiently large n ,

$$\int_{G^c} (\sqrt{n}\|\beta - \beta_0\|)^k \frac{f_{\beta}^{(n)}}{f_{\beta_0}^{(n)}} d\Pi(\beta) < n^{\frac{k}{2}} e^{-nu} \int_{G^c} \|\beta - \beta_0\|^k \pi(\beta) d\beta. \quad (16)$$

By Assumption 1b, the integral on the right hand side is finite, and hence goes to zero as $n \rightarrow \infty$. Along with the fact that for any $\epsilon > 0$, $e^{n\epsilon} \int \frac{f_{\beta}^{(n)}}{f_{\beta_0}^{(n)}} d\Pi(\beta) \rightarrow \infty$ a.s. $[P]$, it will then follow that

$$E \left[\int_{G^c} (\sqrt{n}\|\beta - \beta_0\|)^k d\Pi_n(\beta) \right] \rightarrow 0. \quad (17)$$

Now, for any $\epsilon_1 > 0$ (to be chosen later), using arguments leading up to the proof of Theorem 1 of Sriram et al. (2013), there exists $0 < d < 1$ and some positive constants C'_1, C'_2 such that

$$\begin{aligned} E \left[\int_{G \cap \{\|\beta - \beta_0\| > \epsilon_1\}} (\sqrt{n}\|\beta - \beta_0\|)^k d\Pi_n(\beta) \right] &\leq n^{\frac{k}{2}} M_0^k E [\Pi_n(G \cap \{\|\beta - \beta_0\| > \epsilon_1\})] \\ &\leq n^{\frac{k}{2}} M_0^k E \left[(\Pi_n(G \cap \{\|\beta - \beta_0\| > \epsilon_1\}))^d \right] \leq C'_1 n^{\frac{k}{2}} M_0^k e^{-n\epsilon_1^2 C'_2} \end{aligned}$$

Since the right hand side of the above inequality converges to zero as $n \rightarrow \infty$, we get

$$E \left[\int_{G \cap \{\|\beta - \beta_0\| > \epsilon_1\}} (\sqrt{n}\|\beta - \beta_0\|)^k d\Pi_n(\beta) \right] \rightarrow 0. \quad (18)$$

By Assumption 2 and Lemma 1 of Sriram et al. (2013), there exists a constant $C_3 > 0$ such that

$$\forall \beta : \|\beta - \beta_0\| < \epsilon_1 := \frac{1}{4C_3}, \text{ we have } \left| \log \frac{f_{\beta}}{f_{\beta_0}} \right| \leq C_3 \|\beta - \beta_0\| < \frac{1}{2}.$$

Since for $t < 1/2$, $e^t - 1 < \frac{t}{1-t} < 2t$, and using the fact that $E \left[\log \frac{f_{\beta_0}}{f_{\beta}} \right] > 0$, we have

$$\forall \beta \in \{\|\beta - \beta_0\| < \epsilon_1\}, E \left[\frac{f_{\beta}}{f_{\beta_0}} \right] < 1 + 2E \left[\log \frac{f_{\beta}}{f_{\beta_0}} \right] < e^{-2E \left[\log \frac{f_{\beta_0}}{f_{\beta}} \right]}.$$

We can write $\{\Delta_n < \|\beta - \beta_0\| \leq \epsilon_1\} \subseteq \cup_{j \geq 1} A_{jn}$, where $A_{jn} = \{\beta : j\Delta_n < \|\beta - \beta_0\| \leq (j+1)\Delta_n\}$.

We note using Lemma 2(c) of Sriram et al. (2013) and Assumption 3b that for sufficiently large n , for some constant $C_4 > 0$,

$$\sum_{i=1}^n E \left[\log \frac{f_{i\beta_0}}{f_{i\beta}} \right] \geq C_4 \frac{nj^2 \Delta_n^2}{4}, \forall \beta \in A_{jn}$$

It follows that

$$\begin{aligned} & E \left[\int_{\Delta_n < \|\beta - \beta_0\| \leq \epsilon_1} (\sqrt{n}\|\beta - \beta_0\|)^k \prod_{i=1}^n \frac{f_{i\beta}}{f_{i\beta_0}} \pi(\beta) d\beta \right] \\ & \leq \sum_{j \geq 1} \int_{A_{jn}} (\sqrt{n}\|\beta - \beta_0\|)^k e^{-2 \sum_{i=1}^n E \left[\log \frac{f_{i\beta_0}}{f_{i\beta}} \right]} \pi(\beta) d\beta \\ & \leq (n\Delta_n^2)^{\frac{k}{2}} \sum_{j \geq 1} (j+1)^{\frac{k}{2}} e^{-C_4 \frac{nj^2 \Delta_n^2}{4}} \Pi(A_{jn}) \\ & \leq C_5 \cdot (n\Delta_n^2)^{\frac{k}{2}} e^{-C_4 \frac{n\Delta_n^2}{4}} \cdot \Delta_n^p \sum_{j \geq 1} (j+1)^{\frac{k}{2}+d} e^{-C_4 \frac{n(j^2-1)\Delta_n^2}{4}}. \end{aligned}$$

The last expression uses the fact that $\beta \in \mathbb{R}^p$ and hence $\Pi(A_{jn})$ is less than a constant multiple (say C_5) of $((j+1)\Delta_n^p)$. It is easy to see that the summation in the above expression is bounded for all n . Hence, for some constant C_6 and sufficiently large n , we will have

$$E \left[\int_{\Delta_n < \|\beta - \beta_0\| \leq \epsilon_1} (\sqrt{n}\|\beta - \beta_0\|)^k \prod_{i=1}^n \frac{f_{i\beta}}{f_{i\beta_0}} \pi(\beta) d\beta \right] \leq C_6 \cdot (n\Delta_n^2)^{\frac{k}{2}} e^{-C_4 \frac{n\Delta_n^2}{4}} \cdot \Delta_n^p. \quad (19)$$

For some $C'_4 > 0$ (to be chosen later) let $B_n := \{\beta : \|\beta - \beta_0\| < C'_4 \Delta_n^2\}$. Then, using a similar argument as in the proof of Lemma 3.2 of Kleijn and van der Vaart (2012), we get $\Pi(B_n) \geq K \Delta_n^p$, for some constant K . Hence,

$$\int_{\mathbb{R}^p} \prod_{i=1}^n \frac{f_{i\beta}}{f_{i\beta_0}} \pi(\beta) d\beta \geq \int_{B_n} \prod_{i=1}^n \frac{f_{i\beta}}{f_{i\beta_0}} \pi(\beta) d\beta \geq K \Delta_n^p e^{-nC'_4 \Delta_n^2}. \quad (20)$$

Choosing $C'_4 = C_4/2$, equations (19) and (20) together imply

$$E \left[\int_{\Delta_n < \|\beta - \beta_0\| \leq \epsilon_1} (\sqrt{n} \|\beta - \beta_0\|)^k d\Pi_n(\beta) \right] \leq C_6/K \cdot (n\Delta_n^2)^{\frac{k}{2}} e^{-C_4 \frac{n\Delta_n^2}{8}} \rightarrow 0 \text{ (as } n \rightarrow \infty). \quad (21)$$

Equations (17), (18) and (21) together give the result

$$E \left[\int_{\|\beta - \beta_0\| > \Delta_n} (\sqrt{n} \|\beta - \beta_0\|)^k d\Pi_n(\beta) \right] \rightarrow 0.$$

□

Proof of Lemma 2. Suppose for a vector \mathbf{L} , let $\theta := \mathbf{L}^T \beta$, $\theta_0 := \mathbf{L}^T \beta_0$, $\theta_n^M := \mathbf{L}^T \beta_n^M$ and $\tilde{\theta}_n := \mathbf{L}^T \tilde{\beta}_n$. By abuse of notation, we will continue to use $\Pi_n()$ to denote the posterior distribution of θ from Step 1. Since θ is just a linear combination of β , Lemma 1 implies (for $k \in \{0, 1, 2\}$)

$$\int_{\{\theta: \sqrt{n}|\theta - \theta_0| > \frac{M_n}{2}\}} (\sqrt{n}|\theta - \theta_0|)^k d\Pi_n(\theta) \rightarrow 0 \text{ in probability } [P]. \quad (22)$$

Note that

$$\begin{aligned} & \int_{\{\theta: \sqrt{n}|\theta - \theta_0| > \frac{M_n}{2}\}} n|\theta - \theta_n^M|^2 d\Pi_n(\theta) \\ & \leq 2 \int_{\{\theta: \sqrt{n}|\theta - \theta_0| > \frac{M_n}{2}\}} n|\theta - \theta_0|^2 d\Pi_n(\theta) + 2n|\theta_n^M - \theta_0|^2 \cdot \Pi_n(\sqrt{n}|\theta - \theta_0| > \frac{M_n}{2}) \end{aligned} \quad (23)$$

Using equation (22) and the fact $n|\theta_n^M - \theta_0|^2$ is bounded in probability, we get that the right hand side of equation (23) converges to zero in probability. Hence,

$$\int_{\{\theta: \sqrt{n}|\theta - \theta_0| > \frac{M_n}{2}\}} n|\theta - \theta_n^M|^2 d\Pi_n(\theta) \rightarrow 0 \text{ in probability } [P] \quad (24)$$

Further,

$$\begin{aligned} & \int_{\{\theta: \sqrt{n}|\theta - \theta_n^M| > M_n\}} n|\theta - \theta_n^M|^2 d\Pi_n(\theta) \leq \int_{\{\theta: \sqrt{n}|\theta - \theta_0| > \frac{M_n}{2}\}} n|\theta - \theta_n^M|^2 d\Pi_n(\theta) \\ & \quad + \int_{\{\theta: \sqrt{n}|\theta - \theta_0| \leq \frac{M_n}{2}, \sqrt{n}|\theta - \theta_n^M| > M_n\}} n|\theta - \theta_n^M|^2 d\Pi_n(\theta) \end{aligned} \quad (25)$$

First term on right hand side of (25) is same as that in equation (24). Since $\sqrt{n}|\theta - \theta_0| \leq \frac{M_n}{2}$ and $\sqrt{n}|\theta - \theta_n^M| > M_n$ would imply $\sqrt{n}|\theta_n^M - \theta_0| > \frac{M_n}{2}$, for any $\epsilon > 0$, for the second term, we can write

$$P \left(\int_{\{\theta: \sqrt{n}|\theta - \theta_0| \leq \frac{M_n}{2}, \sqrt{n}|\theta - \theta_n^M| > M_n\}} n|\theta - \theta_n^M|^2 d\Pi_n(\theta) > \epsilon \right) \leq P(\sqrt{n}|\theta_n^M - \theta_0| > \frac{M_n}{2})$$

Since the right hand side goes to zero as $n \rightarrow \infty$, the second term of equation (25) converges to zero in probability. Therefore, we have established that

$$\int_{\{\theta: \sqrt{n}|\theta - \theta_n^M| > M_n\}} n|\theta - \theta_n^M|^2 d\Pi_n(\theta) \rightarrow 0 \text{ in probability } [P] \quad (26)$$

Now, let Q_n denote the distribution of $\sqrt{n}(\theta - \theta_n^M)$ under $\theta \sim \Pi_n$. Let Q denote $N(0, \mathbf{L}^T V^{-1} \mathbf{L})$. Then Theorem 1 implies that Q_n converges to Q with respect to the total variation norm. Therefore,

$$\text{for any fixed } M, \int_{|x| \leq M} x^2 dQ_n \rightarrow \int_{|x| \leq M} x^2 dQ \text{ as } n \rightarrow \infty, \text{ and } \int_{|x| \leq M} x^2 dQ \rightarrow \int x^2 dQ \text{ as } M \rightarrow \infty.$$

So, there exists a sequence $\{M_n\}$ such that $\int_{|x| \leq M_n} x^2 dQ_n \rightarrow \int x^2 dQ$ as $n \rightarrow \infty$. For such a sequence, using equation 26, we get $\int x^2 dQ_n \rightarrow \int x^2 dQ$ as $n \rightarrow \infty$. Equivalently,

$$\int n(\theta - \theta_n^M)^2 d\Pi_n \rightarrow \mathbf{L} V^{-1} \mathbf{L} \text{ as } n \rightarrow \infty. \quad (27)$$

Following the same approach as above, we can also conclude that:

$$\int |n(\theta - \theta_n^M)| d\Pi_n \rightarrow \int |x| dQ(x) \text{ as } n \rightarrow \infty. \quad (28)$$

$$\int (n(\theta - \theta_n^M))^+ d\Pi_n \rightarrow \int x^+ dQ(x) \text{ as } n \rightarrow \infty. \quad (29)$$

It is now easy to see that equations (28) and (29) imply that

$$\sqrt{n}(\mathbf{L}^T \boldsymbol{\beta}_n^M - \mathbf{L}^T \tilde{\boldsymbol{\beta}}_n) \rightarrow 0 \text{ in probability } [P]$$

This along with equation (27) implies that

$$\int n(\theta - \tilde{\theta}_n)^2 d\Pi_n \rightarrow \mathbf{L} V^{-1} \mathbf{L} \text{ as } n \rightarrow \infty.$$

Since the above two results hold for any vector \mathbf{L} , statements (a) and (b) of Lemma 2 follow. \square

Proof of Lemma 3. Recall $Z_i = Y_i - \mathbf{X}_i^T \beta_0$. Let $P_i(\cdot)$ denote the cumulative distribution function for the p.d.f $p_i(\cdot)$. In the lines of proof of Theorem 4.1 in Koenker (2005), we write

$$U_n(\delta) := \log \frac{f_{\beta_0 + \frac{\delta}{\sqrt{n}}}^{(n)}}{f_{\beta_0}^{(n)}} = U_{1n}(\delta) + E[U_{2n}] + (U_{2n}(\delta) - E[U_{2n}(\delta)])$$

where,

$$U_{1n}(\delta) := -\delta^T \frac{1}{\sqrt{n}} \sum_{i=1}^n (\tau - I_{\{Z_i \leq 0\}}) \mathbf{X}_i, \text{ and}$$

$$U_{2n}(\delta) := \sum_{i=1}^n \int_0^{\frac{\mathbf{X}_i^T \delta}{\sqrt{n}}} (I_{\{Z_i \leq s\}} - I_{\{Z_i \leq 0\}}) ds = \sum_{i=1}^n U_{2ni}(\delta).$$

Further, by Taylor's formula,

$$E(U_{2n}(\delta)) = \sum_{i=1}^n \int_0^{\frac{\mathbf{X}_i^T \delta}{\sqrt{n}}} (P_i(\mathbf{X}_i^T \beta_0 + s) - P_i(\mathbf{X}_i^T \beta_0)) ds = \sum_{i=1}^n \int_0^{\frac{\mathbf{X}_i^T \delta}{\sqrt{n}}} p_i(\mathbf{X}_i^T \beta_0 + r_{in}(s)) \cdot s \, ds,$$

where $0 \leq r_{in}(s) \leq s \leq \left| \frac{\mathbf{X}_i^T \delta}{\sqrt{n}} \right|$. Therefore, by Assumptions 2 and 5, we have

$$\begin{aligned} & \left| E(U_{2n}(\delta)) - \frac{1}{2n} \sum_{i=1}^n p_i(\mathbf{X}_i^T \beta_0) \delta^T \mathbf{X}_i \mathbf{X}_i^T \delta \right| \\ & \leq \sum_{i=1}^n \int_0^{\frac{\mathbf{X}_i^T \delta}{\sqrt{n}}} C |r_{in}(s)|^\eta \cdot s \, ds \leq C \left(\max_{i \geq 1} \frac{|\mathbf{X}_i^T \delta|}{\sqrt{n}} \right)^\eta \cdot \delta^T \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^T \delta. \end{aligned}$$

It follows that,

$$\sup_{\delta \in K} \left| E(U_{2n}(\delta)) - \frac{1}{2n} \sum_{i=1}^n p_i(\mathbf{X}_i^T \beta_0) \delta^T \mathbf{X}_i \mathbf{X}_i^T \delta \right| \rightarrow 0 \text{ in probability } [P]. \quad (30)$$

To complete the proof, it is enough to show $\sup_{\delta \in K} |U_{2n}(\delta) - E[U_{2n}(\delta)]| \rightarrow 0$ in probability $[P]$. To show this, let $\delta_i = \arg \max_{\delta \in K} |U_{2ni}(\delta) - E[U_{2ni}(\delta)]|$. δ_i is possibly random since it can depend on Z_i , but is well defined since $U_{2ni}(\delta)$ is a continuous (random) function on a compact set K . Then,

$$\begin{aligned} & P \left(\sup_{\delta \in K} |U_{2n}(\delta) - E[U_{2n}(\delta)]| > \epsilon \right) \leq P \left(\sum_{i=1}^n \sup_{\delta \in K} |U_{2ni}(\delta) - E[U_{2ni}(\delta)]| > \epsilon \right) \\ & = P \left(\sum_{i=1}^n |U_{2ni}(\delta_i) - E[U_{2ni}(\delta_i)]| > \epsilon \right) \leq \sum_{i=1}^n \frac{E(U_{2ni}^2(\delta_i))}{\epsilon^2}. \end{aligned} \quad (31)$$

The last step uses Chebyhev's inequality. Further, let $G = \sup_{\delta \in K} \sup_{i \geq 1} |\mathbf{X}_i^T \delta|$. In particular,

$|\mathbf{X}_i^T \boldsymbol{\delta}_i| \leq G$. Without loss of generality, if $\mathbf{X}_i^T \boldsymbol{\delta} > 0$, $\forall i$, the right hand side of Equation (31) is

$$\begin{aligned} & \leq \sum_{i=1}^n \frac{E \left(\int_0^{\frac{\mathbf{X}_i^T \boldsymbol{\delta}_i}{\sqrt{n}}} (I_{\{Z_i \leq s\}} - I_{\{Z_i \leq 0\}}) ds \right)^2}{\epsilon^2} \leq \sum_{i=1}^n \frac{E \left(\int_0^{\frac{G}{\sqrt{n}}} (I_{\{Z_i \leq s\}} - I_{\{Z_i \leq 0\}}) ds \right)^2}{\epsilon^2} \\ & \leq \sum_{i=1}^n \frac{G^2}{n} \frac{E \left(I_{\{Z_i \leq \frac{G}{\sqrt{n}}\}} - I_{\{Z_i \leq 0\}} \right)^2}{\epsilon^2} \leq \frac{G^2}{n \epsilon^2} \sum_{i=1}^n \left(P_i(\mathbf{X}_i^T \boldsymbol{\beta}_0 + \frac{G}{\sqrt{n}}) - P_i(\mathbf{X}_i^T \boldsymbol{\beta}_0) \right) \\ & \leq \frac{G^2}{n \epsilon^2} \sum_{i=1}^n |p_i(\mathbf{X}_i^T \boldsymbol{\beta}_0 + r_{in})| \frac{G}{\sqrt{n}}, \text{ for some } |r_{in}| \leq \frac{G}{\sqrt{n}} \text{ using Taylor's formula.} \end{aligned}$$

A consequence of assumptions **2** and **5** is that $\{p_i(\mathbf{X}_i^T \boldsymbol{\beta}_0 + r_{in}), i \geq 1\}$ are uniformly bounded. Hence, we can conclude that $P(\sup_{\boldsymbol{\delta} \in K} |U_{2n}(\boldsymbol{\delta}) - E[U_{2n}(\boldsymbol{\delta})]| > \epsilon) \rightarrow 0$. \square

References

- Alhamzawi, R. and Yu, K. (2013), “Conjugate priors and variable selection for Bayesian quantile regression,” *Computational Statistics and Data Analysis*, 64, 209–219.
- Benoit, D. F. and Van den Poel, D. (2012), “Binary Quantile Regression: a Bayesian Approach based on the Asymmetric Laplace Distribution,” *Journal of Applied Econometrics*, 1174–1188.
- Diaconis, P. and Freedman, D. (1986), “On the consistency of Bayes estimates (with discussion).” *The Annals of Statistics*, 14, 1–26.
- Kleijn, B. J. K. and van der Vaart, A. W. (2012), “The Bernstein-Von-Mises theorem under misspecification.” *Electronic Journal of Statistics*, 6, pp. 354–381.
- Koenker, R. (2005), *Quantile Regression (Econometric Society Monographs)*, Cambridge University Press.
- Koenker, R. and Bassett, Gilbert, J. (1978), “Regression Quantiles,” *Econometrica*, 46, 33–50.
- Kozumi, H. and Kobayashi, G. (2011), “Gibbs sampling methods for Bayesian quantile regression,” *Journal of statistical computation and simulation*, 81, 1565–1578.
- Müller, U. K. (2013), “Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix.” *Econometrica*, 81, 1805–1849.
- Sriram, K., Ramamoorthi, R. V., and Ghosh, P. (2013), “Posterior consistency of Bayesian quantile regression based on the misspecified asymmetric Laplace density,” *Bayesian Analysis*, 8, 479–504.
- Waldmann, E., Kneib, T., Yue, Y. R., Lang, S., and Flexeder, C. (2013), “Bayesian semiparametric additive quantile regression,” *Statistical Modelling*, 13, 223–252.
- Yu, K. and Moyeed, R. A. (2001), “Bayesian Quantile Regression,” *Statistics and Probability Letters*, 54, 437–447.

- Yu, K., van Kerm, P., and Zhang, J. (2005), “Bayesian quantile regression: an application to the wage distribution in 1990s Britain,” *Sankhyā*, 57, 359–377.
- Yue, Y. R. and Rue, H. (2011), “Bayesian Inference for Additive Mixed Quantile Regression Models,” *Compututational Statistics and Data Analysis*, 55, 84–96.